

Kahane-Khinchin type Averages

Omer Friedland¹

February 5, 2008

Abstract

We prove a Kahane-Khinchin type result with a few random vectors, which are distributed independently with respect to an arbitrary log-concave probability measure on \mathbb{R}^n . This is an application of small ball estimate and Chernoff's method, that has been recently used in the context of Asymptotic Geometric Analysis in [1], [2].

1 Introduction

The classical **Kahane's inequality** (cf. Kahane [8]) states that for any $1 \leq p < \infty$ there exists a constant $K_p > 0$ such that

$$\int_0^1 \left\| \sum_{i=1}^n r_i x_i \right\| dt \leq \left(\int_0^1 \left\| \sum_{i=1}^n r_i x_i \right\|^p dt \right)^{1/p} \leq K_p \int_0^1 \left\| \sum_{i=1}^n r_i x_i \right\| dt, \quad (1)$$

holds true for every n and arbitrary choice of vectors $x_1, \dots, x_n \in X$, where X is a normed space with the norm $\|\cdot\|$, and r_i is a Rademacher function that is given by $r_i(t) = \text{sign} \sin(2^{i-1}\pi t)$, $i \geq 1$ (for more information about Rademacher functions, see, e.g. Milman and Schechtman [13], section 5.5). Kwapien [9] proved that $K_p \sim \sqrt{p}$ is a constant depending only on p .

Note that, in view of the definition of $r_i(t)$, these integrals in (1) can also be represented as averages (for $p \geq 1$)

$$\int_0^1 \left\| \sum_{i=1}^n r_i(t) x_i \right\|^p dt = \text{Ave}_\pm \left\| \sum_{i=1}^n \pm x_i \right\|^p. \quad (2)$$

Bourgain, Lindenstrauss and Milman [4] proved that, if v_1, \dots, v_n are unit vectors in a normed space $(\mathbb{R}^n, \|\cdot\|)$ then, for any $\varepsilon > 0$ there exists a constant $C(\varepsilon) > 0$, such that $N = C(\varepsilon)n$ random sign vectors $\varepsilon(1), \dots, \varepsilon(N) \in \{-1, 1\}^n$ satisfy with probability greater than $1 - e^{-cn}$ that

$$(1 - \varepsilon) \|x\| \leq \frac{1}{N} \sum_{j=1}^N \left\| \sum_{i=1}^n \varepsilon(j)_i x_i v_i \right\| \leq (1 + \varepsilon) \|x\|, \quad (3)$$

¹omerfrie@post.tau.ac.il; address: School of Mathematical Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel.

for every $x \in \mathbb{R}^n$, where $|||x||| = \text{Ave}_\pm \|\sum_{i=1}^n \pm x_i v_i\|$.

Obviously, the norm $|||\cdot|||$ is an unconditional norm, i.e., it is invariant under the change of signs of the coordinates. As (3) shows, it is sufficient to average $O(n)$ terms in (2) rather than 2^n , in order to obtain a norm that is isometric to $|||\cdot|||$, and in particular symmetrize our original norm $\|\cdot\|$ with selected vectors $\{v_i\}_{i=1}^n$ to become almost unconditional.

In this paper, we change the settings given in (3), in two ways:

First, random sign vectors are replaced by random vectors $a(1), \dots, a(N)$ with an arbitrary log-concave probability measure μ .

Second, we are now interested in using a small number $N = (1 + \delta)n$ of random vectors, at the cost of an isomorphic, instead of an almost isometric, comparison of $|||x||| = \int_{\mathbb{R}^n} \|\sum_{i=1}^n a_i x_i v_i\| d\mu(a)$ with the random norm $|||x|||_N = \frac{1}{N} \sum_{j=1}^N \|\sum_{i=1}^n a(j)_i x_i v_i\|$.

Let v_1, \dots, v_n be unit vectors in a normed space $(\mathbb{R}^n, \|\cdot\|)$. Define a norm $|||\cdot|||$ on \mathbb{R}^n : $|||x||| = \int_{\mathbb{R}^n} \|\sum_{i=1}^n a_i x_i v_i\| d\mu(a)$, where a_i is the i^{th} coordinate of the vector a and μ is a log-concave probability measure μ on \mathbb{R}^n .

Theorem. *Let $N = (1 + \delta)n$, $0 < \delta < 1$, and let $\{a(1), \dots, a(N) \in \mathbb{R}^n\}$ be a set of N independent random vectors, distributed with respect to a log-concave probability measure μ on \mathbb{R}^n . Then*

$$\mathbb{P} \left\{ \forall x \in \mathbb{R}^n \ c(\delta) |||x||| \leq \frac{1}{N} \sum_{j=1}^N \left\| \sum_{i=1}^n a(j)_i x_i v_i \right\| \leq C |||x||| \right\} \geq 1 - e^{-c'n},$$

where $c(\delta) = (c\delta)^{1+\frac{2}{\delta}}$ and $c, c', C > 0$ are universal constants.

Remark. It is easy to see that once you learn the theorem for small δ , it holds for large δ as well. Thus we may always assume that $\delta < \delta_0$ for some universal constant δ_0 .

Remark. The norm $|||\cdot|||$ depends on the choice of vectors the $v_1, \dots, v_n \in \mathbb{R}^n$. Different choice of vectors defines a different norm for $x \in \mathbb{R}^n$. For example, if we choose $v_1 = \dots = v_n$, it reduces the question above to the scalar case, the norm $|||\cdot|||$ is the Euclidean norm and we get a result of isomorphic Khinchin-type inequality; these type of results were proved by Litvak, Pajor, Rudelson, Tomczak-Jaegermann and Vershynin [11],[12] and also by Artstein-Avidan, Friedland and Milman [2].

Remark. The question above was also investigated for any $0 < p < \infty$. See, e.g. Bourgain [3], Rudelson [14], Giannopoulos and Milman [5], Guédon and Rudelson [6]. We shall focus on the case of $p = 1$.

Acknowledgement. I thank my supervisor Vitali Milman for encouragement, support and interest in this work. I thank Sasha Sodin and Shiri Artstein-Avidan for useful discussions and remarks on preliminary versions of this note. I thank Sergey Bobkov for suggesting me to look at the continuous case of this problem, and last but not least, I thank Apostolos Giannopoulos for useful remarks and referring me to the paper of Latała [10].

2 Proof of Theorem

Before we proceed, we give a short description of Chernoff's method. The following lemma, which is a version of Chernoff's bounds, gives estimates for the probability that at least βN trials out of N succeed, when the probability of success in one trial is p (cf. Hagerup and Rub [7]).

Lemma 1 (Chernoff). *Let Z_1, \dots, Z_N be independent Bernoulli random variables with mean $0 < p < 1$, that is, Z_i takes value 1 with probability p and value 0 with probability $(1 - p)$. Then we have*

$$\begin{aligned} 1) \quad & \mathbb{P}\{Z_1 + \dots + Z_N \geq \beta N\} \geq 1 - e^{-NI(\beta, p)} \quad \text{when } \beta < p, \\ 2) \quad & \mathbb{P}\{Z_1 + \dots + Z_N \geq \beta N\} \leq e^{-NI(\beta, p)} \quad \text{when } \beta > p, \end{aligned}$$

where $I(\beta, p) = \beta \ln \frac{\beta}{p} + (1 - \beta) \ln \frac{1 - \beta}{1 - p}$.

In the questions above, essentially, we are looking for upper and lower bounds of $\frac{1}{N} \sum_{j=1}^N \|\sum_{i=1}^n a(j)_i x_i v_i\|$. Upper bounds are relatively easy to obtain, and quite often do not require new methods, but only the use of large deviation inequalities like Bernstein's inequality and some net argument. Obtaining lower bounds is different, usually one needs small ball probabilities, which are hard to get, and some extra delicate arguments which are closely related to the context of the question at hand. Here comes Chernoff's method, if one has a small ball probability for one trial, using Chernoff's bounds, the estimate of the average of many trials can be amplified. For more detailed description of this method, see Artstein-Avidan, Friedland and Milman [1], [2].

Finally, let us analyze the function

$$I(\beta, p) = \beta \ln \frac{\beta}{p} + (1 - \beta) \ln \frac{1 - \beta}{1 - p} = u(\beta) - \beta \ln p - (1 - \beta) \ln (1 - p),$$

where we denoted $u(\beta) = \beta \ln \beta + (1 - \beta) \ln (1 - \beta)$. The term $u(\beta)$ is a negative, convex function which approaches 0 as $\beta \rightarrow 0$ and as $\beta \rightarrow 1$, and is symmetric about $1/2$ where it has a minima equal to $-\ln 2$. Thus the whole exponent in Lemma 1 is of the form

$$e^{-NI(\beta, p)} = p^{\beta N} (1 - p)^{(1 - \beta)N} e^{-Nu(\beta)} \leq (1 - p)^{(1 - \beta)N} 2^N. \quad (4)$$

Proof of Theorem. We estimate the following probability

$$\mathbb{P}\left\{a(1), \dots, a(N) : \forall x \in S_{||\cdot||}^{n-1} \quad c(\delta) \leq |||x|||_N \leq C\right\},$$

where $|||x|||_N = \frac{1}{N} \sum_{j=1}^N \|\sum_{i=1}^n a(j)_i x_i v_i\|$ and $S_{||\cdot||}^{n-1} = \{x \in \mathbb{R}^n : |||x||| = 1\}$.

Remark. The norm $|||x|||_N$ is a *random* norm depending on the choice of N random vectors $a(1), \dots, a(N)$.

It is clear that the probability above is greater than the following one

$$\begin{aligned} 1 & - \mathbb{P}\{\exists x \in S_{||\cdot||}^{n-1}, |||x|||_N > C\} \\ & - \mathbb{P}\{(\forall y \in S_{||\cdot||}^{n-1}, |||y|||_N \leq C) \text{ and } (\exists x \in S_{||\cdot||}^{n-1}, |||x|||_N < c(\delta))\}. \end{aligned}$$

Upper bound: We begin by estimating the first term

$$\mathbb{P}\{\exists x \in S_{|||\cdot|||}^{n-1}, |||x|||_N > C\} .$$

This is relatively easy, and does not require a new method; we do it in a similar way to the one in [4]: let $\mathcal{N} = \{y(i)\}_{i=1}^m$ be a $\frac{1}{2}$ -net with respect to $|||\cdot|||$ on $S_{|||\cdot|||}^{n-1}$, it is known that such a net exists with $m \leq 5^n$. For each $1 \leq i \leq m$ we consider the random variables $\{X_{i,j}\}_{j=1}^N$ defined by

$$X_{i,j} = \left\| \sum_{k=1}^n a(j)_k y(i)_k v_k \right\| - r ,$$

where $r = \mathbb{E} \left\| \sum_{k=1}^n a_k x_k v_k \right\| = 1$ and $y(i)_k$ is the k^{th} coordinate of the vector $y(i)$. Clearly, for each i and j , $X_{i,j}$ has mean 0 and $\|X_{i,j}\|_{\psi_1} \leq b$ for some absolute constant $b > 0$ (see Milman and Schechtman [13], App. III). Now, using the well-known Bernstein's inequality which we shall use in the form of ψ_1 estimate (see Vaart and Wellner [15]):

Lemma 2 (Bernstein). *Let Y_1, \dots, Y_N be independent random variables with mean 0 such that for some $b > 0$ and every i , $\|Y_i\|_{\psi_1} \leq b$. Then, for any $t > 0$,*

$$\mathbb{P}\left\{\left|\frac{1}{N} \sum_{i=1}^N Y_i\right| > t\right\} \leq 2 \exp(-cN \min(\frac{t}{b}, \frac{t^2}{b^2})) ,$$

where $c > 0$ is an absolute constant.

we deduce that for any $t > 0$ and every $1 \leq i \leq m$, we have

$$\mathbb{P}\{a(1), \dots, a(N) : \left|\frac{1}{N} \sum_{j=1}^N X_{i,j}\right| > t\} \leq 2 \exp(-cN \min(\frac{t}{b}, \frac{t^2}{b^2})) ,$$

which implies that for a point $y(i) \in \mathcal{N}$ and for any $t > 1$ we have

$$\begin{aligned} \mathbb{P}\{a(1), \dots, a(N) : \frac{1}{N} \sum_{j=1}^N \left\| \sum_{k=1}^n a(j)_k y(i)_k v_k \right\| > t\} &\leq \\ &\leq 2 \exp(-cN \min(\frac{t-1}{b}, \frac{(t-1)^2}{b^2})). \end{aligned}$$

The obvious way to make this probability small enough to handle a large net is to increase t , and obviously we shall get a worse upper bound constant. So, we choose t such that

$$2e^{-cN \min(\frac{t-1}{b}, \frac{(t-1)^2}{b^2})} \cdot 5^n \leq e^{-n} ,$$

for example $t = 6 \max(\frac{b}{\sqrt{c}}, \frac{b}{c}) + 1$. Then, with probability at least $1 - e^{-n}$, for every $1 \leq i \leq m$ we have

$$\frac{1}{N} \sum_{j=1}^N \left\| \sum_{k=1}^n a(j)_k y(i)_k v_k \right\| \leq t .$$

We thus have an upper bound for a net on the sphere. It is standard to transform this to an upper bound estimate on *all* the sphere (this is an important difference between lower and upper bounds). One uses a consecutive approximation of a point on the sphere by points from the net to get that $|||x|||_N \leq 2t = 2t$ for every $x \in S_{|||\cdot|||}^{n-1}$. This completes the proof of the upper bound, where $C = 12 \max(\frac{b}{\sqrt{c}}, \frac{b}{c}) + 2$ is a universal constant.

Lower bound: We now turn to estimate the second term

$$\mathbb{P}\{(\forall y \in S_{|||\cdot|||}^{n-1}, |||y|||_N \leq C) \text{ and } (\exists x \in S_{|||\cdot|||}^{n-1}, |||x|||_N < c(\delta))\}. \quad (5)$$

Note that when estimating this term, we know *in advance* that the (random) norm $|||\cdot|||_N$ is bounded from above on the sphere $S_{|||\cdot|||}^{n-1}$ (i.e. $\forall y \in S_{|||\cdot|||}^{n-1}$ we have $|||y|||_N \leq C$, where C comes from the upper bound). This is crucial to transform a lower bound on a net on the sphere to a lower bound on the whole sphere. For the lower bound we use Chernoff's method, as described above, to estimate the probability in (5).

Let us denote by p the probability that for a random vector $a \in \mathbb{R}^n$ we have $\|\sum_{i=1}^n a_i x_i v_i\| \geq \alpha$, where $\alpha > 0$ and x is some point on $S_{|||\cdot|||}^{n-1}$:

$$p = \mu(\{a \in \mathbb{R}^n : \|\sum_{i=1}^n a_i x_i v_i\| \geq \alpha\}).$$

If "doing an experiment" means checking whether $\|\sum_{i=1}^n a_i x_i v_i\| \geq \alpha$ (where $a \in \mathbb{R}^n$ is a random vector) then for $|||x|||_N$ to be greater than some c , it is enough that $(c/\alpha)N$ of the experiments succeed.

Of course, we will eventually not want to do this on *all* points x on the sphere, but just on some dense enough set. So, first we estimate the probability p :

Lemma 3. *There exists a universal constant $\gamma > 0$ such that for any $x \in S_{|||\cdot|||}^{n-1}$, we have*

$$\mu(\{a \in \mathbb{R}^n : \|\sum_{i=1}^n a_i x_i v_i\| \leq \gamma\}) \leq \frac{2}{3}.$$

Proof of Lemma 3. Let us define $A_x = \{a \in \mathbb{R}^n : \|\sum_{i=1}^n a_i x_i v_i\| \leq \gamma_x\}$ for $x \in S_{|||\cdot|||}^{n-1}$, then A_x is convex and symmetric set. We take $\gamma_x > 0$ to be the number such that $\mu(A_x) = \frac{2}{3}$. Applying Borell's lemma (see Milman and Schechtman [13], App. III.3) we get, for all $t > 1$,

$$\mu(\{a \in \mathbb{R}^n : \|\sum_{i=1}^n a_i x_i v_i\| > \gamma_x t\}) \leq \frac{2}{3} \left(\frac{1}{2}\right)^{(t+1)/2},$$

and consequently,

$$\begin{aligned}
1 = |||x||| &= \int_{\mathbb{R}^n} \left\| \sum_{i=1}^n a_i x_i v_i \right\| d\mu(a) = \int_0^\infty \mu\left(\left\| \sum_{i=1}^n a_i x_i v_i \right\| > t\right) dt \\
&\leq \gamma_x + \int_{\gamma_x}^\infty \mu\left(\left\| \sum_{i=1}^n a_i x_i v_i \right\| > t\right) dt \\
&\leq \gamma_x + \gamma_x \int_1^\infty \frac{2}{3} \left(\frac{1}{2}\right)^{\frac{t+1}{2}} dt \\
&\leq \gamma_x (1 + c_4)
\end{aligned}$$

for some universal constant $c_4 > 0$ which doesn't depend on x . Therefore, we get that $\gamma_x \geq \frac{1}{1+c_4} > 0$. Now, we take $\gamma = \frac{1}{1+c_4}$. For this $\gamma > 0$ and any $x \in S_{|||\cdot|||}^{n-1}$ we have $\mu(\{a \in \mathbb{R}^n : \left\| \sum_{i=1}^n a_i x_i v_i \right\| \leq \gamma\}) \leq \frac{2}{3}$ \square

Now, we can use the following lemma of Latała with the set

$$C_x = \{a \in \mathbb{R}^n : \left\| \sum_{i=1}^n a_i x_i v_i \right\| \leq \gamma\},$$

for which we proved above that $\mu(C_x) \leq \frac{2}{3} = b < 1$.

Lemma 4 (Latała [10]). *For each $b < 1$ there exists a constant $c_b > 0$ such that for every log-concave probability measure μ and every measurable convex, symmetric set C with $\mu(C) \leq b$ we have*

$$\mu(tC) \leq c_b t \mu(C) \quad \text{for } t \in [0, 1].$$

Notice that, for *any* point x , we can make the probability as small as we like by reducing t . This allows us to use a simple net: take θ -net \mathcal{N} in $S_{|||\cdot|||}^{n-1}$, with less than $(\frac{3}{\theta})^n$ points.

For every $x \in S_{|||\cdot|||}^{n-1}$ there is a vector $y \in \mathcal{N}$ such that $|||x - y||| \leq \theta$, and we have $|||y|||_N \leq |||x|||_N + |||x - y|||_N \leq c + C\theta$ (where $c = c(\delta)$ and C comes from the upper bound). Therefore, we bound (5) by

$$\mathbb{P}\{\exists y \in \mathcal{N} : |||y|||_N \leq c + C\theta\}. \quad (6)$$

By Lemma 4, for a given $y \in \mathcal{N}$ we have for any $0 < t < 1$ that

$$p = \mu(\{a \in \mathbb{R}^n : \left\| \sum_{i=1}^n a_i y_i v_i \right\| \geq t\gamma\}) \geq 1 - c_1 t,$$

where $c_1 = c_b \frac{2}{3}$.

We return to our scheme, in order to estimate the probability in (6), assume that $\beta t \gamma \geq c + C\theta$, where β shall be the portion of good trials out of N , and t another constant that we choose later such that $p > \beta$. So, we know that for

$\beta < 1 - c_1 t$ (which is hardly a restriction, t will be very small and so will β), from Lemma 1 for a given $y \in \mathcal{N}$ we have

$$\mathbb{P}\{\|y\|_N \geq \beta t \gamma\} \geq \mathbb{P}\{Z_1 + \dots + Z_N \geq \beta N\} \geq 1 - e^{-NI(\beta, p)},$$

where $\mathbb{P}\{Z_i = 1\} = \mu(\{a \in \mathbb{R}^n : \|\sum_{i=1}^n a_i y_i v_i\| \geq t \gamma\})$.

We choose β so that $(1 + \delta)(1 - \beta) = 1 + \frac{\delta}{2}$, hence $\beta = \frac{\delta}{2(1+\delta)}$. We choose $\theta = \beta t \gamma / 2C$, where C comes from the upper bound. To make sure that the probability above holds for all points in the net we ask that

$$\begin{aligned} e^{-NI(\beta, p)} \cdot |\mathcal{N}| &\leq 2^N (1 - p)^{(1-\beta)N} \cdot \left(\frac{3}{\theta}\right)^n \leq \\ &\leq 2^N (c_1 t)^{(1+\frac{\delta}{2})n} \cdot \left(\frac{3}{\theta}\right)^n \leq \frac{1}{2} e^{-c'n}. \end{aligned}$$

For the first inequality we use (4), we choose $t = (c_2 \delta)^{\frac{2}{\delta}}$, for some universal constant $c_2 > 0$, and get the lower bound for each point of the net \mathcal{N} .

Now using the upper bound, for every $x \in S_{||\cdot||}^{n-1}$ there is a vector $y \in \mathcal{N}$ such that $\|x - y\| \leq \theta$, therefore we have

$$\|x\|_N \geq \|y\|_N - \|x - y\|_N \geq \beta t \gamma - \theta C =: c,$$

$c = c(\delta) = (c_3 \delta)^{1+\frac{2}{\delta}}$, where $c_3 > 0$ is an absolute constant. Thus the proof of the lower bound, and of the Theorem, is completed. \square

References

- [1] S. Artstein-Avidan, O. Friedland, V.D. Milman. *Geometric applications of Chernoff-type estimates and a zigzag approximation for balls*. Proc. Amer. Math. Soc. 134 (2006), no. 6, 1735–1742.
- [2] S. Artstein-Avidan, O. Friedland, V.D. Milman. *Geometric Applications of Chernoff-Type Estimates*, Geometric Aspects of Functional Analysis, Israel Seminar 2004–2005, Lecture Notes in Mathematics, Vol. 1910.
- [3] J. Bourgain. *Random points in isotropic convex sets*. Convex Geometric Analysis (Berkeley, CA, 1996), 53–58, Math. Sci. Res. Inst. Publ., 34, Cambridge Univ. Press, Cambridge, 1999.
- [4] J. Bourgain, J. Lindenstrauss, V.D. Milman. *Minkowski sums and symmetrizations*. Geometric aspects of functional analysis (1986/87), Lecture Notes in Math., 1317, Springer, Berlin, 1988, 44–66.
- [5] A.A. Giannopoulos, V.D. Milman. *Concentration property on probability spaces*. Adv. Math. 156 (2000), no. 1, 77–106.
- [6] O. Guédon, M. Rudelson. *L_p -moments of random vectors via majorizing measures*. Adv. Math. 208 (2007), no. 2, 798–823.

- [7] T. Hagerup, C. Rub. *A guided tour of Chernoff bounds*. Info Proc Lett 33 (1990) no. 6, 305-308.
- [8] J. P. Kahane. *Some random series of functions*. second edition, Cambridge Studies in Advanced Math., 5, (1985), Cambridge University press.
- [9] S. Kwapień. *A theorem on the Rademacher series with vector valued coefficients*. Probability in Banach spaces, Oberwolfach 1975, Lecture Notes in Mathematics, Vol. 526, Springer-Verlag, Berlin-Heidelberg-New York, 1976, pp. 157-158.
- [10] R. Latała. *On the equivalence between geometric and arithmetic means for log-concave measures*. in Convex geometric analysis (Berkeley, CA, 1996), 123-127, Math. Sci. Res. Inst. Publ. 34, Cambridge Univ. Press, Cambridge 1999.
- [11] A.E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann. *Smallest singular value of random matrices and geometry of random polytopes*. Adv. Math. 195 (2005), no. 2, 491–523.
- [12] A.E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, R. Vershynin. *Random Euclidean embeddings in spaces of bounded volume ratio*. C. R. Math. Acad. Sci. Paris 339 (2004), no. 1, 33–38.
- [13] V.D. Milman and G. Schechtman. *Asymptotic theory of finite dimensional normed spaces*. Springer Lecture Notes, 1200, (1986).
- [14] M. Rudelson. *Random vectors in the isotropic position*. J. Funct. Anal. 164 (1999), no. 1, 60–72.
- [15] A.W.v.d. Vaart, J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics, 2000.